GENERAL INFORMATION

Title of Dataset: Census of the Ecosystem of Decentralized Autonomous Organizations

Author/Principal Investigator Information

Name: Andrea Peña Calvin
ORCID: 0009-0004-1317-6088
Institution: Dept. Software Engineering and Artificial Intelligence, Universidad Complutense de Madrid, Spain
Email: andpen03@ucm.es

Author/Associate or Co-investigator Information

Name: Andrew Schwartz
ORCID: 0009-0002-8913-2562
Institution: University of Colorado Boulder, USA
Email: andrew.t.schwartz@colorado.edu

Author/Associate or Co-investigator Information

Name: Javier Arroyo
ORCID: 0000-0001-6127-7538
Institution: Institute of Knowledge Technology, Universidad Complutense de Madrid, Spain
Email: jarroyog@ucm.es

Author/Associate or Co-investigator Information

Name: Samer Hassan
ORCID: 0000-0002-4696-779X
Institution: Institute of Knowledge Technology, Universidad Complutense de Madrid, Spain and
Berkman Klein Center for Internet and Society, Harvard University, Cambridge, MA, USA
Email: shassan@cyber.harvard.edu

Date of data collection: Data was collected from July 20 to August 8, 2023.

SHARING/ACCESS INFORMATION

Licenses placed on the data: CC BY 4.0

Links to publications that cite or use the data: https://doi.org/10.1145/3589335.3651481

Was data derived from another source? Yes. More information on the publication.

Recommended citation for this dataset:

Peña-Calvin, A., Schwartz, A., Arroyo, J., & Hassan, S. (2024). Concentration of Power and Participation in Online Governance: the Ecosystem of Decentralized Autonomous Organizations. *Companion Proceedings of the ACM Web Conference, 13–17, 2024, Singapore,* doi: https://doi.org/10.1145/3589335.3651481

METHODOLOGICAL INFORMATION

The description of the methods used for the generation of data, for processing it and the quality-assurance procedures performed on the data can be found here: https://doi.org/10.1145/3589335.3651481

DESCRIPTION

The dataset comprises three CSV files: `deployments.csv`, `proposals.csv`, and `votes.csv`, each containing essential information regarding DAOs deployments, their proposals, and the corresponding votes. It includes data from various Decentralized Autonomous Organizations (DAOs) platforms, namely Aragon, DAOHaus, DAOstack, Realms, Snapshot and Tally.

The file `deployments.csv` provides insights into the general aspects of DAO deployments, including the platform it is deployed in, the number of proposals, unique voters, votes cast, and estimated voting power.

In the `proposals.csv` file, comprehensive information about all proposals associated with the deployments is contained, including its date, the number of votes it received, and the total voting power voters employed on that proposal.

In `votes.csv`, data regarding the votes cast for the deployment proposals is recorded. It includes the blockchain address of the voter, the weight of the vote in voting power, and the day it was cast.

KEYWORDS

Blockchain, DAOs, Decentralized Autonomous Organizations, Decentralized governance, Decentralized Web, Decision making, Online Communities, Online Governance, Proposals, Votes, Voting power, Web3.
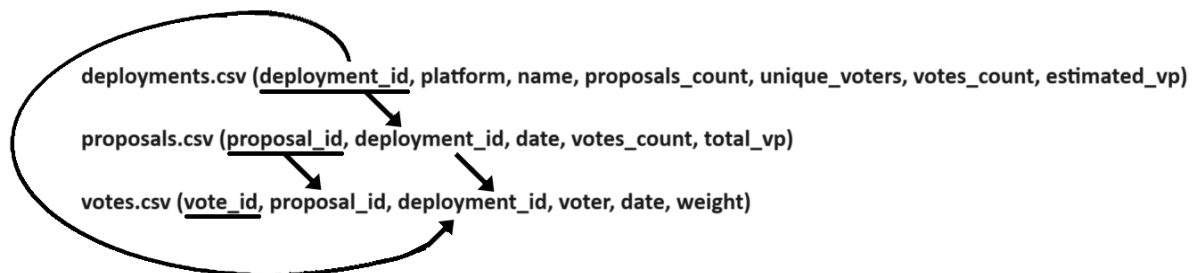
FILE OVERVIEW

File List:

- deployments.csv: contains the information about the general aspects of each DAO deployment.
- proposals.csv: contains the information about all the proposals of the deployments.
- votes.csv: contains the information about the votes of the deployments proposals.

VISUAL GUIDE FOR UNDERSTANDING THE DATASET

Unique identifiers of each file are underlined, and arrows illustrate how the different files relate to each other (for users with more experience with databases, the underlined variable represents the primary key of the table, while arrows depict the relationships with foreign keys).

DATA-SPECIFIC INFORMATION FOR: [deployments.csv]

Number of variables: 7

Number of cases/rows: 10,541

Variable List:

- **deployment_id**: unique deployment identifier.
- **platform**: platform to which the deployment belongs.
- **name**: name of the DAO to which the deployment belongs.
- **proposals_count**: number of proposals of the deployment.
- **unique_voters**: number of different voters that have ever participated in the deployment proposals.
- **votes_count**: number of votes cast by deployment voters.
- **estimated_vp**: estimated voting power of the deployment.

Missing data codes: NaN.

DATA-SPECIFIC INFORMATION FOR: [proposals.csv]

Number of variables: 5

Number of cases/rows: 186,228

Variable List:

- **proposal_id**: unique proposal identifier.
- **deployment_id**: deployment to which the proposal belongs.
- **date**: date of the proposal
- **votes_count**: number of votes cast in the proposal.
- **total_vp**: sum of the voting power of the votes received

Missing data codes: there is no missing data.

DATA-SPECIFIC INFORMATION FOR: [votes.csv]

Number of variables: 6

Number of cases/rows: 21,606,463

Variable List:

- **vote_id**: unique vote identifier.
- **proposal_id**: proposal to which the vote belongs.
- **deployment_id**: deployment to which the vote belongs.
- **voter**: address of the voter that cast the vote.
- **date**: date when the vote was cast.
- **weight**: voting power of the vote.

Missing data codes: NaN.